PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : H04H 5/00

A1

(11) International Publication Number: WO 97/37449

(43) International Publication Date: 9 October 1997 (09.10.97)

(21) International Application Number: PCT/US97/05141

(22) International Filing Date: 28 March 1997 (28.03.97)

(30) Priority Data:
08/627,947 3 April 1996 (03.04.96) US

(71) Applicant: COMMAND AUDIO CORPORATION [US/US]; Suite 310, 805 Veterans Boulevard, Redwood City, CA 94083 (US).

(72) Inventor: MORRISON, Eric, F.; 3980 Jefferson Avenue, Redwood City, CA 94060 (US).

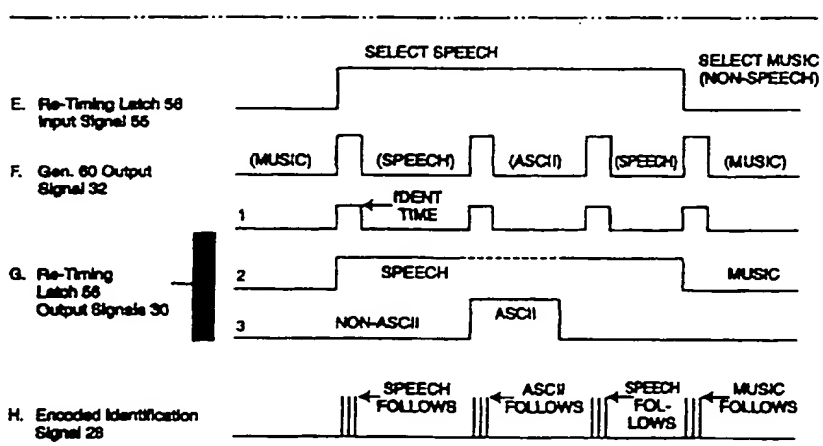
(74) Agents: KLIVANS, Norman, R. et al.; Skjerven, Morrill, MacPherson, Franklin & Friel, Suite 700, 25 Metro Drive, San Jose, CA 95110 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: DIGITAL AUDIO DATA TRANSMISSION SYSTEM BASED ON THE INFORMATION CONTENT OF AN AUDIO SIGNAL



(57) Abstract

The data rate of speech and non-speech audio is selectively reduced by respective compression techniques based upon the information content of the type of signal. A composite audio information signal formed of speech and non-speech audio is applied to both a voice encoder and a wide-band audio compression encoder. An audio-type detection circuit examines the speech spectrum as well as the entire frequency spectrum and dynamic range of the audio information and generates a selection signal indicating whether the signal is speech or non-speech audio. A composite encoded audio signal is produced by intermingling the outputs of the encoders in response to the selection signal. The composite encoded audio signal and an identification signal indicative of the audio signal type are transmitted to respective receivers at the reduced data rates for storage, and subsequent decoding and retrieval by a listener as an audible signal in response to the transmitted identification signal.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

DIGITAL AUDIO DATA TRANSMISSION SYSTEM BASED ON THE INFORMATION CONTENT OF AN AUDIO SIGNAL

CROSS REFERENCE TO RELATED PATENT

This invention is related to a commonly assigned U.S. Patent 5,406,626, issued April 11, 1995 to John O. Ryan entitled Radio Receiver for Information Dissemination Using Subcarrier, and to copending U.S. Patent Applications Serial No. 08/181,394 filed January 12, 1994, to John O. Ryan entitled A Method and System for Audio Information Dissemenation Using Various Modes of Operation, and Serial No. 08/223,641 filed April 6, 1994 to John O. Ryan entitled A Method and System for Information Dissemenation Using Various Modes of Transmission.

BACKGROUND OF THE INVENTION

The invention relates to the transmission of digital audio signals over narrow band data channels and, more particularly, to the reduction of the data rate of transmission and reception of a digital audio signal based on the information content of the signal, that is, based on whether the audio signal is speech or non-speech. The channels consist of point-to-point digital telephony links and audio broadcast services where normally narrow bandwidth channels would degrade the quality of the recovered audio signals.

A digitized audio source signal requires considerable channel bandwidth to transmit the full frequency range and dynamic range of the original analog source signal. Digital audio compression techniques, such as proposed for the Moving Picture Experts Group-2 (MPEG-2) transmissions described in the industry standard ISO 11172-3, take advantage of the psycho-acoustical characteristics of the ear-brain combination to reduce the channel bandwidth by reducing the data rate of the digitized signal. In a practical application of the concept, the reductions achieved generally are insufficient when compared to the bandwidth of the original analog source signal.

Voice encoders used for transmitting digitized speech in extremely narrow bandwidths find application in the telecommunications industry where only narrow

bandwidth channels are available. The encoder reduces the data rate of the speech signals by converting the information using a model of the human voice generation process. The coefficients of the model representing a measurement of the speaker's voice are transmitted to a receiver which converts the coefficients to a voice presentation of the original source signal. Such a technique provides exceptional data rate compression of spoken audio, but only is applicable to speech signals since it is based on recognition and electronic modeling of speech. It follows that these voice encoders work very efficiently for voice signals but are unable to process other types of non-speech signals such as music.

Accordingly, in order to transmit and receive both speech and non-speech signals such as music, it is necessary to provide an alternate data compression scheme when such non-speech audio signals are to be transmitted and received. Thus, in any practical audio signal transmission/ reception system where both speech and non-speech are intermingled to form the audio information, some means must be provided to detect the type of audio signal and to adapt the compression scheme to the audio type, whereby the technique used to compress the respective audio signal may be optimized to maximize the data rate while providing the best possible speech and non-speech quality.

SUMMARY OF THE INVENTION

The invention circumvents the problems associated with optimizing the data rate of speech and non-speech audio information while maintaining the best quality possible for each type of audio in applications where the signals are intermingled. To this end, the invention reduces the data rate of the digital audio signal based on the information content of the signal. The type of signal to be data compressed (usually speech or music) is determined and the optimum compression, based on information content, is applied.

Advantageously, the reduced data rate requires less channel bandwidth and/or allows more signals on a given transmission channel. In the case of a system where the received audio information is stored in a memory for later retrieval, the information may be sent at a higher speed thereby reducing the transmission time as well.

The majority of communicated information is in the form of the spoken word by a recognizable voice. In order to optimize the efficiency of transmitting audio information, significant reductions in data rate are achieved by applying the digitized

speech signal to a voice encoder (vocoder). For example, a typical vocoder operating on a typical 64 kbit/sec source signal can convert the signal to a data rate of 2 4 kbit/sec, a coding gain of 27 times.

In the present invention, a complex audio information signal (combinations of speech and music) is applied to both a vocoder and a conventional full range audio compression encoder, using an audio-type selection technique that examines the speech spectrum as well as the entire frequency spectrum and dynamic range of the audio information for subsequent selectable compression. To this end, the high coding gain speech vocoder is used to compress the speech signals and the full range encoder with a lower coding gain is used to compress the composite signal that includes speech, music and other non-speech signals. An audio-type detection circuit is used to measure the audio input signal and to decide if the signal is speech or non-speech. In one embodiment, the detection circuit monitors the speech frequency spectrum and measures the occurrence of pauses indicative of a speech signal. The detection circuit also measures the energy content outside the speech range of frequencies. A combination of the results of these measurements determines if the audio information is speech or non-speech. In an alternative embodiment, the internal signal processing within the vocoder is used to provide an external signal indicative of which type of audio signal is present. If the signal is speech the low data rate vocoder path is selected in response to a selection signal, and if it is non-speech the higher data rate compression encoder path is selected. In addition, an identification signal is generated to identify the type of audio data signal that is present.

The encoded composite audio signal is transmitted along with the identification signal, for reception by suitable receivers which include respective memories for storing the composite audio and identification signal for subsequent retrieval. Upon retrieval, the respective audio signals are separated and decoded in response to the identification signal, whereby the original speech and non-speech signals are made available to a listener in the form of an audible signal.

Another form of information signal suitable for conversion to audio is ASCII text which may be selected for transmission to data receivers along with the two other types of audio data signals and a unique identification signal. The identification signal comprises a code which identifies the type of signal selected, and is multiplexed with

the digitized encoded audio information for transmission. The code subsequently directs the selection of the desired decoder in the data receivers

A typical system for encoding, transmitting, receiving and decoding audio signals is described in the patent and applications of previous mention, that is, U.S. Patent 5,406,626 and USSN 08/181,394 and 08/223,641, the descriptions of which are herein incorporated by reference in their entirety.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1A AND 1B is a block diagram illustrating an encoder system environment for encoding and transmitting audio information, in which the invention decision making detector means may be utilized.

FIGURE 2A AND 2B is a block schematic diagram illustrating one embodiment of the decision making detector means of the present invention.

FIGURE 3 is a block diagram illustrating a decoder system environment for receiving the encoded and transmitted audio information in accordance with the decoding means of the invention.

FIGURE 4A AND 4BA-4H is a timing diagram illustrating the respective waveforms appearing at various inputs and outputs of the circuit components shown in FIGURE 2A AND 2B.

FIGURE 5 is a block diagram illustrating an alternative embodiment of the decision making detector means of the invention.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

FIGURE 1A AND 1B depicts an encoder system 10 which comprises the invention environment, wherein digitized audio information, hereinafter referred to as a digital audio source signal, is supplied on a lead 12 in either serial or parallel format and is sample rate converted by a sample rate converter circuit 14 to produce a 64 kbit/sec data signal. The data signal is applied to a vocoder 16 The sampling rate and dynamic range of the digital audio source signal on the input lead 12 to the encoder system will usually be greater than the 64 kbit/sec digitized audio signal required by the vocoder 16 Thus, prior to the vocoder 16 the signal is sample rate converted from the source rate to 64 kbit/sec

via the sample rate converter circuit 14. Typical data rates for the encoder system 10 are shown in FIGURE 1A AND 1B.

The vocoder 16 is of the type used in the telecommunications industry such as the voice codec IMBE™ manufactured by Digital Voice Systems, Inc., Burlington, Massachusetts.

The audio source signal on lead 12 also is applied via a compensating delay 20 to a wide-band digital audio compression encoder 18 such as those used for transmitting entertainment programming in compressed form such as, for example, digital audio broadcast transmissions. Typical of a wide-band audio compression encoder is the MUSICAM® encoder manufactured by Philips. This type of audio compression is described as Audio Layer II in the ISO 11172-3 standard for audio sub-band coding. The audio source signal 12 further is applied to an audio-type decision making detector 22 of the invention, further described in FIGURE 2A AND 2B. The vocoder processing delay can be of the order of hundreds of milliseconds, hence the compensating delay 20 is inserted ahead of the audio compression encoder to maintain time coincidence at the outputs of the components 16, 18. The outputs of components 16, 18, 22 are in turn coupled to the inputs of a data selector/multiplexer 24.

The efficiency of a digital compression system is expressed as coding gain (CG) and is given by CG = input data rate/output data rate. A vocoder (such as 16) producing a 2.4 kbit/sec output for a 64 kbit/second input typically has a coding gain of 26.67. Audio compression encoders (such as 18) typically have coding gains of the order of 8 to 16 depending on the signal quality level desired.

A second input to the encoder system is a digital ASCII text signal on a lead 26 of the order of 100 bit/sec that, following transmission, is converted to pseudo audio information signals by a receiver such as described below in FIGURE 3 using a method of a text-to-speech converter such as BeSTspeech™ manufactured by Berkeley Speech Technologies of Berkeley, California. The ASCII text is treated as a separate audio information signal and is applied to a buffer at the input of the audio-type detector 22, further described in FIGURE 2A AND 2B. Selection between digital audio source signal 12 and ASCII text signal 26 is performed as data from each source becomes available. The ASCII text signal is the third input to the digital data selector and

multiplexer 24. Reading of the ASCII signal and inclusion in the data path uses conventional data processing techniques.

Selection between the vocoder 16 and the audio compression encoder 18 is made by the audio-type decision making detector 22 based on measurement of the incoming digital audio source signal as described below in FIGURE 2A AND 2B. The precise timing of the selection between the encoders 16, 18 is initiated at common block boundaries of the two digital audio-type signals as further described below. The detector 22 provides an audio-type identification signal via a lead 28, a selection signal via a bus 30 and a re-timed ASCII text via a lead 34, to the data selector/multiplexer 24. A block timing signal is supplied via a lead 32 from the detector 22 to the vocoder 16 and encoder 18. Signal 32 controls the boundary timing of the blocks of data generated by the encoders 16, 18. The data selector/multiplexer 24 includes a multiplexing circuit for supplying an intermingled composite digital audio/identification output signal which includes the audio-type identification signal. The output signal is supplied via a lead 36 to a conventional transmission system (depicted at 38) for transmission in typical fashion to a decoder system of respective multiple audio receiver means, an example of which is further depicted in FIGURE 3. The audio/identification output signal may be in parallel or serial digital format.

By way of operation in general, the decision making detector 22 of FIGURE 1A AND 1B looks at the energy in the frequency spectrum covering the range of speech of the audio source signal on bus 12, and measures the length, in time, of the typical pauses of silence occurring between syllables. The detector 22 further measures the energy content outside the voice range of frequencies. A combination of the results of the two detections determines if the audio is speech or is other non-speech sounds such as music. From this determination a selection signal is generated on bus 30 and is used to control the data selector/multiplexer 24 which intermingles the speech and non-speech signals into the composite audio output signal The selection signal is formed of three timing signals on respective leads of the bus 30, as further described in FIGURE 4A AND 4B. The intermingled selection signal first is re-timed via a re-timing latch (FIGURE 2A AND 2B) to cause the switching between types of audio to occur at the phase synchronous block boundaries of the corresponding audio signals being encoded in the

audio compression encoder 18 and vocoder 16

The data identification signal is generated on the lead 28 and is unique to each type of audio signal, that is, speech, non-speech and ASCII, and is multiplexed with the selected audio signals via the data selector/multiplexer 24 to provide the composite audio/identification output signal on lead 36. The identification signal is used subsequently as a control signal for a complementary demultiplexer in the audio receiver means (FIGURE 3).

The encoder system of FIGURE 1A AND 1B also determines the time of insertion of ASCII text by examining the occupancy of an internal buffer memory in the ASCII data path, further described in FIGURE 2A AND 2B. The selection signal from this measurement also is re-timed to occur on the block boundaries of the audio signals being processed in the encoders 16, 18. The combined selection signals operate the data selector/ multiplexer 24 to provide the composite audio/identification output signal on the lead 36, which thus includes the identification signal on lead 28 multiplexed with the audio data. The ASCII text signal is re-timed by the re-timing latch of previous mention for inclusion with the other audio data in response to a buffer occupancy signal shown in FIGURE 2A AND 2B.

Referring now to FIGURE 2A AND 2B, the audio-type decision making detector 22 of the invention is shown in greater detail. The digitized audio source signal is supplied in either a serial or parallel format via the lead 12 to an automatic gain control circuit (AGC) 40, and thence to a band-pass filter (BPF) 42 of a first identification (ident) path 43. The audio source signal also is applied to a delay network 41 and thence to a non-inverting input of a subtractor circuit 44 of a second ident path 45. The delay network 41 compensates for the delay introduced by the band-pass filter 42 so that the signals appearing on leads 39 and 47, comprising the input signals to the subtractor circuit 44, are in time with each other. The output of the BPF 42 is supplied to a pause detector circuit 46 (described later) as well as to an inverting input of the subtractor circuit 44. The output of the pause detector circuit 46 is supplied to an AND gate 48 and the output of the subtractor circuit 44 is supplied to a threshold circuit 50 and thence to a second input of the AND gate 48. A reference signal which determines the operating threshold is coupled to the threshold circuit 50 via a lead 52. The logic output of the AND gate 48 is

coupled to a hysteresis circuit 54 and thence via a lead 55 to a re-timing latch 56 as an initial selection signal. The output of the re-timing latch 56 is the selection signal of previous mention on bus 30. The output of the hysteresis circuit 54 also is supplied via the lead 55 to a timing generator 60 to re-time the selection process by making it occur at the common block boundaries of the compressed audio data signals. The re-timed selection signal appears on the bus 30.

The pause detector 46 looks for short pauses between bursts of data indicating typical speech. A pause is defined as a significant reduction in the instantaneous level of the audio signal with respect to the average audio level occurring for a period of 50 to 150 milliseconds and at a rate of 1 to 3 times per second. The precise timings are determined empirically and vary depending on the speed of the speech and the language spoken. If a string of pauses meeting the above or similar criteria is met over a period of time, the pause detector produces a logic one at its output, lead 49. If pauses are not detected, the output is a logic zero.

The ASCII text on lead 26 is supplied to an ASCII buffer 58 which supplies a buffer occupancy signal via a lead 59 to the timing generator 60, to the re-timing latch 56 and to an identification code latch 62 whose output is the identification signal of previous mention on the lead 28. The output of the buffer 58 is supplied on the lead 34 as the re-timed ASCII text signal of previous description. A timing signal from the timing generator 60 is the block timing signal on the lead 32, which also is supplied to the re-timing latch 56 and the identification code latch 62 as well as to the encoders 16, 18 of FIGURE 1A AND 1B.

Regarding more particularly the operation of FIGURE 2A AND 2B, the digitized audio source signal is applied to the AGC 40 to maintain a fixed output level for all audio input levels. Following the AGC, the audio is applied to the speech band-pass filter BPF 42 covering the frequency range from 300 Hz to 3 kHz, which represents the frequency band containing the maximum speech energy. Unlike other types of sounds, speech consists of syllables and pauses, whereby detection of the pauses is one indication of a speech signal. Accordingly, the pause detector circuit 46 provides a logic one output if a relatively large number of pauses are measured in a unit of time, indicating a speech signal. If the pause detector circuit 46 does not detect a given large number of pauses in

the signal, the circuit 46 outputs a logic zero. The logic signal is applied as one input to the logic AND gate 48.

The band-pass signal from the BPF 42 is subtracted from the flat frequency response signal supplied by the AGC 40 via the subtractor circuit 44 to produce a non-speech signal representing frequency components outside the range of normal speech. This signal is applied to the threshold circuit 50 which produces a logic one output if the audio level is below a predetermined threshold set by the reference level on the lead 52. A logic zero output is produced if the audio level is greater than the threshold, indicating that the signal is a non-speech signal such as music. The logic signal from threshold circuit 50 is the second input to the AND function.

In accordance with the invention, if pauses are detected in the limited bandwidth signal of path 43 and sufficient energy is not present in the remaining range of frequencies, that is, in the non-speech signal in the path 45, the output of the AND gate 48 is a logic one, indicating a speech signal is present with no other sounds of significant level.

The truth table below illustrates in further detail the output states of the pause detector circuit 46, the threshold circuit 50, the AND gate 48 as well as the encoder selection, for possible combinations of input conditions.

condition	pause detector 46	threshold circuit 50	AND gate 48	selection
wide-band audio (non-speech/ music)	X	0	0	audio compression encoder 18
pauses in audio, wide-band audio present (non-speech/ music)	1	0	0	audio compression encoder 18
pauses in audio, narrow band audio present (speech)	1	1	1	vocoder 16
no audio present, or very long pauses (no signal)	1	1	1	vocoder 16

Hysteresis is applied to the AND logic output signal by the circuit 54 to prevent the signal from toggling in the range of uncertainty. The logic signal further is re-timed by the re-timing latch 56 of previous mention to align it with the common block boundaries of the two types of encoded audio of the encoder outputs, in response to the timing generator 60.

The ASCII text information on the lead 26 is written to the ASCII buffer 58 and the buffer occupancy of the buffer 58 is constantly monitored. As the buffer reaches the full state the internal fullness measurement initiates a buffer nearly full signal and the buffer 58 supplies a pause signal, that is, the buffer occupancy signal, on lead 59 to the timing generator 60, to the re-timing latch 56 and to the identification code latch 62. The buffer is read out at a high data rate, relative to the ASCII input signal on lead 26. The audio encoders 16, 18 of FIGURE 1A AND 1B are instructed via the block timing signal 32 to store their converted audio data temporarily while the ASCII text data is transferred from the ASCII buffer 58 to the transmission path 34. When the ASCII buffer empties, the buffer fullness measurement function disables the ASCII read process and the encoders 16, 18 are enabled to continue outputting their respective audio signals to the data selector/multiplexer 24. The latter circuit 24 multiplexes the two audio signals of speech and non-speech into a composite audio signal in response to the selection signal on the bus 30. The identification signal on the lead 28 also is multiplexed into the composite audio signal to provide the composite audio/identification output signal on the lead 36 for transmission in conventional fashion

via the transmission system indicated at 38.

FIGURE 4A AND 4BA-4H illustrates further the operation of the decision making detector 22 in the course of determining the type of audio information supplied on the input lead 12. To this end, when the ASCII buffer 58 is nearly full, the buffer occupancy signal on lead 59 goes to a high binary state as shown in FIGURE 4A AND 4BA. The output 32 of the timing generator 60 supplies the block timing signal indicative of the boundaries of the blocks of data generated for the vocoder 16 and audio compression encoder 18, as shown in FIGURE 4A AND 4BC. At the trailing edge of the transition of the block boundary signal following the buffer occupancy signal 59 (FIGURE 4A AND 4BA), the ASCII buffer 58 is read using an internal read signal shown in FIGURE 4A AND 4BB. During this period of time the data of both thevocoder 16 and audio compression encoder 18 are temporarily stored as depicted via the dimension line 64 in FIGURE 4A AND 4B. The read and re-timed ASCII text information is depicted in FIGURE 4A AND 4BD. When the buffer 58 empties, the buffer occupancy signal on lead 59 transitions to a low state as shown in FIGURE 4A AND 4BA.

The timing signal indicative of the selection of speech (vocoder 16) or non-speech (encoder 18) is supplied to the re-timing latch 56 from thehysteresis circuit 54 via the lead 55, and is shown in FIGURE 4A AND 4BE. The latch 56 also receives the occupancy signal on lead 59 which indicates the selection of ASCII text (FIGURE 4A AND 4BA). The third input to the re-timing latch 56 is the block timing signal on lead 32 which indicates the boundaries of the audio-type signals and the type of signal to be selected, that is, speech or non-speech. The signal 32 is depicted in FIGURE 4A AND 4BF which corresponds to thewaveform of FIGURE 4A AND 4BC. The output of the re-timing latch 56 comprises the selection signal on the bus 30 which includes three timing signals shown in FIGURE G1, G2, G3.

Signal GI of the selection signal indicates the time for selection of the identification code signal on lead 28 by the data selector/multiplexer 24. Signal G2 indicates the time for the selection of the speech signal from thevocoder 16, or the non-speech signal from the audio compression encoder 18. Signal G3 indicates the time for the selection of the ASCII text by the data selector/multiplexer 24.

The identification code latch 62 receives the block timing signal on lead 32 indicating block boundaries andvocoder 16 or audio compression encoder 18 modes, and the buffer occupancy signal on lead 59 indicating the selection of ASCII text information. The identification code signal from the latch 62 on lead 28 is multiplexed with the data via the data selector/multiplexer 24 in response to the signal G1, as previously described. The coded identification signal is depicted in FIGURE 4A AND 4BH and is timed to occur within the

corresponding time periods of the block timing signal on lead 32 of FIGURE 4A AND 4BC and 4F

Referring now to FIGURE 3, the transmitted composite audio/identification signal is supplied to a memory 66 integral with a decoder system 70 of the receiver means of previous mention. The stored audio then may be recovered when desired by a user in response to a user control signal on a lead 67 The recovered audio and identification signals are supplied via a lead 72 to an identification decoder 68 of the decoder system 70. The memory 66 and decoder system 70 comprise the receiver means for receiving and utilizing a restored version of the digital audio source signal originally supplied to the encoder system 10 of FIGURES 1, 2. Such a receiver means is discussed in the patent andcopending applications of previous reference. The identification decoder 68 searches for and separates the identification signal from the composite audio/identification signal. The identification signal as previously discussed indicates, in time, when a change occurs in the type of audio signal. The identification decoder 68 detects the unique codes that identify the type of audio data received by the input 72 from the memory 66. The decoded identification signal is supplied via a lead 76 to a cross-fade switch 78 as a control signal. The composite audio signal is supplied via a lead 80 to avocoder decoder 82 and also to a wide-band audio decompression decoder 84. Thevocoder decoder 82 extracts the speech signal from the composite audio signal and supplies it to a speech input of the cross-fade switch 78. The wide-band decoder 84 extracts the non-speech signal from the composite audio signal and supplies it to a non-speech input of the switch 78 via a compensating delay 86, which compensates for the decoder 82 signal processing time. The cross-fade switch 78 generally is conventional in function and, in response to the controlling identification signal on lead 76, provides a soft switching of the speech and non-speech signals to produce a resulting smoothly intermingled digital audio output signal on an output bus 88. The audio output signal corresponds to the digital audio source signal originally supplied via the bus 12 to the encoder system 10 of FIGURES 1, 2. The digital audio signal on output bus 88 is converted to analog format whereby the audio information may beransduced via a conventional amplifier/speaker system (not shown) into a signal for aural presentation to a listener.

Although the invention has been described herein relative to specific embodiments, various additional features and advantages will be apparent from the description and drawings. For example, a vocoder (that is, vocoder 16) also may be used to detect the presence of speech or non-speech signals as an alternate to a corresponding portion of the audio-type decision making detector 22. The vocoder measures the frequency components of speech usually using a fasfourier transform or other frequency selective transform. If thevocoder produces an accurate electrical representation of the incoming signal with the normal speech bandwidth as evidenced by comparing the reconstructed voice coded signal with the input signal in the frequency domain, then a safe

assumption can be made that the input signal in question is a voice coded signal. If the comparison shows significant differences exist between the two compared signals, then a safe assumption can be made that the signal is a non-speech or music signal. The resulting signal of such a comparison may be applied to the hysteresis function. 54 of FIGURE 2A AND 2B in place of the components 40-48 of the decision making detector 22.

FIGURE 5 depicts the use of avocoder 16' as the alternative of previous mention for making the audio-type decision indicative of whether the audio signal is speech or non-speech. To this end, the sample rate converted audio signals of 64kbits are supplied to the vocoder 16' which then provides an output on a lead 90 indicative of the accuracy of the incoming signal relative to the normal speech bandwidth, and thus indicative of whether a speech signal is present. The output on lead 90 is compared with the threshold reference level on lead 52 via the threshold circuit 50. The threshold circuit provides the selection signal on lead 55 as a logic one if the audio level is below the threshold level indicating a speech signal. A logic zero output is provided if the audio level is greater than the threshold level which provides a selection signal on lead 55 indicating a non-speech signal.

Thus the scope of the invention is intended to be defined by the following claims and their equivalents.

What is claimed is:

1. Apparatus for encoding digital audio information formed of audio signals such as speech signals and non-speech signals, comprising:

means for generating a selection signal indicative of the speech signal or the non-speech signal;

means responsive to the selection signal for providing an identification signal indicative of the audio signals for inclusion with the selected audio signals; and

means for selectively intermingling the speech signal, the non-speech signal and the identification signal in response to the selection signal.

2. The apparatus of claim 1 wherein the generating means includes:

means for detecting whether the information is a speech signal or a non-speech signal; and

said generating means being responsive to the detecting means.

3. The apparatus of claim 2 wherein the detecting means includes:

first means for generating a first signal indicative of the presence or absence of a speech signal;

second means for generating a second signal indicative of the presence or absence of the non-speech signal; and

logic means for generating said selection signal in response to the first and second signals.

4. The apparatus of claim 3 wherein the first signal is representative of a preselected ratio of pauses in the audio information to indicate the presence or absence of the speech signal.

5. The apparatus of claim 4 where the first means includes:

a filter for passing apassband signal in a frequency range which contains the maximum speech energy; and

a pause detector responsive to the filter for generating a logic state indicative of an occurrence of successive pauses in the audio information.

6 The apparatus of claim 5 wherein the second means includes:

means responsive to the passband signal and the audio information for providing a third signal representing frequency components outside the range of the speech signal; and

means responsive to the third signal and to a predetermined threshold level for producing a logic state indicative of the level of energy in the third signal.

7. The apparatus of claim 6 wherein the producing means includes an audio level threshold circuit for comparing the third signal with the predetermined threshold level.

8. The apparatus of claim 6 wherein the logic means includes AND logic responsive to the logic states of the pause detector and the producing means, for generating said selection signal.

9. The apparatus of claim 8 further including voice encoder means for encoding the speech signal;

wherein the logic state of the pause detector is a first state, the logic state of the threshold means is a first state, and the selection signal from the AND logic is a first state indicative of the presence of the speech signal; and

wherein the voice encoder means is selected in response to the first state of the selection signal.

10. The apparatus of claim 8 further including wide-band audio compression encoder means for encoding the non-speech signal;

wherein the logic states of the pause detector and of the threshold means are unlike, and the selection signal from the AND logic is a second state indicative of the presence of a non-speech signal; and

wherein the wide-band encoder means is selected in response to the second state of the selection signal.

11. The apparatus of claim 2 further including:

voice encoder means for encoding the speech signal;

wide-band audio compression encoder means for encoding the non-speech signal; and

the intermingling means includes multiplexer means receiving the encoded speech and non-speech signals and the identification signal for intermingling the signals in response to the

selection signal.

12. The apparatus of claims 2 wherein said means for providing includes.

timing generator means responsive to the selection signal for synchronizing the identification signal with the occurrence of the audio signals; and

latch means responsive to the timing generator means for providing the identification signal.

13. The apparatus of claim 12 wherein the audio signals include an ASCII text signal, including:

buffer means for selectively supplying the ASCII text signal; and

said timing generator means being responsive to the buffer means for storing the speech and non-speech signals in response to the buffer means supplying the ASCII text signal.

14. The apparatus of claim 2 wherein the detecting means includes:

voice encoder means for receiving and compressing the audio signals;

means for comparing the accuracy of the reconstructed voice coded signal with the audio signals; and

said means for generating including means for generating the selection signal indicative of a speech signal in response to an accurate comparison and indicative of a non-speech signal in response to significant inaccuracy in the comparison.

15. The apparatus of claim 14 wherein the means for comparing includes a threshold circuit.

16. Apparatus for transmitting and receiving digital audio information including speech and non-speech signals, comprising:

means for detecting whether the information is a speech signal or a non-speech signal and for generating a selection signal indicative thereof;

means responsive to the selection signal for providing an identification signal indicative of the type of audio information;

means for selecting the speech signal, the non-speech signal or the identification signal for transmission in response to said selection signal,

means for separating the identifying signal upon receiving the transmitted

information: and

means for intermingling the speech signal and non-speech signal subsequent to the receiving in response to said separated identifying signal, to restore the digital audio information.

17. The apparatus of claim 16 including:

means for transmitting and receiving the identifying signal together with the speech and non-speech signals; and

means integral with the receiving means for storing the received speech, non-speech and identifying signals for subsequent recovery.

18. The apparatus of claim 17 further including:

means for encoding the speech signal and the non-speech signal with respective optimum compression based on the energy content of each signal; and

wherein the selecting means selects the encoded speech, the non-speech or the identification signal for transmission in response to said selection signal.

19. The apparatus of claim 18 wherein:

said receiving means includes decoder means for separating the speech signal and the non-speech signal; and

switching means responsive to the separated identifying signal for combining the speech and non-speech signals into an intermingled analog signal corresponding to a restoration of the digital audio information, for audible presentation.

20. The apparatus of claim 19 wherein:

said encoding means includes a narrow band speech encoder and a wide-band non-speech encoder; and

said decoding means includes a narrow band speech decoder and a wide-band non-speech decoder.

21. Apparatus for reducing the transmission data rate of digital audio information formed of speech signals and non-speech signals, comprising:

means for detecting whether the information is a speech or a non-speech signal and for generating a selection signal indicative thereof;

means for separately encoding the speech and non-speech signals with respective optimum compression based on the information energy content of the signals;

means responsive to the detecting and generating means for producing a signal identifying the speech signal and the non-speech signal; and

means for intermingling the encoded speech signal and the encoded non-speech signal in response to the selection signal, for transmission at said reduced data rate

22. The apparatus of claim 21 wherein the detecting means includes:

means for generating a first signal indicative of the occurrence of a large number of pauses in a unit of time in a selected frequency range of the audio information corresponding to a speech signal; and

means for generating a second signal indicative of audio frequency components outside the selected frequency range corresponding to a non-speech signal.

23. The apparatus of claim 22 wherein the generating means includes:

logic means for producing in response to the first and second signals a logic state identifying the presence of a speech signal or a non-speech signal.

24. The apparatus of claim 23 wherein the first signal generating means includes:

a filter for providing apassband signal of said selected frequency range; and

a pause detector responsive to thepassband signal for generating a logic state corresponding to said first signal.

25. The apparatus of claim 24 wherein:

said filter provides apassband in a frequency range of maximum speech energy; and

said logic means is an AND gate.

26. The apparatus of claim 22 wherein the second signal generating means includes:

summing means responsive to thepassband signal and the audio information for providing a third signal representing audio frequency components outside the selected frequency range; and

threshold means responsive to the third signal for providing a logic state corresponding to said second signal.

27. The apparatus of claim 26 wherein:

said summing means is asubtractor for subtracting thepassband signal from the audio information; and

said threshold means includes a threshold input of a selected audio level for comparison to the third signal.

28. The apparatus of claim 21 wherein:

the encoding means includes a voicecoder for encoding the speech signal and a wide-band audio compression encoder for encoding the non-speech signal; and

the intermingling means includes a selector/multiplexer circuit for selecting the encoded speech signal, the encoded non-speech signal or the identifying signal in response to the selection signal.

29. The apparatus of claim 28 including:

means for transmitting the encoded speech and non-speech signals selected by the selector/multiplexer circuit along with the identifying signal; and

receiver means receiving the transmitted encoded speech and non-speech signals for selectively decoding in response to the identifying signal the respective audio signals into a reassembled audio signal corresponding to the digital audio information, for audible presentation.

30. The apparatus of claim 29 wherein the receiver means includes:

memory means for temporarily storing the transmitted signals;

means coupled to the memory means for separating the identifying signal from the encoded speech and non-speech signals;

decoder means for separately decoding each of the encoded speech and non-speech signals; and

switching means for selecting the decoded speech or the non-speech signal in response to the separated identifying signal to form the reassembled audio signal for audible presentation.

31. A method for reducing the transmission rate of digital audio information formed of speech signals and non-speech signals, comprising the steps of:

detecting whether the audio information is the speech signal or the non-speech signal;

encoding the speech signal in a respective narrow frequency range;

encoding the non-speech signal in a respective wide-band frequency range outside of the narrow frequency range;

generating in response to the detecting step a selection signal indicative of the speech signal and the non-speech signal; and

selecting the encoded speech signal or the encoded non-speech signal for transmission at the reduced rate in response to the selection signal.

32. The method of claim 31 wherein the step of detecting includes the steps of:

detecting if the audio information contains a relatively large succession of pauses indicative of a speech signal; and

generating a first logic signal indicative of whether the signal is or is not the speech signal.

33. The method of claim 32 wherein the step of detecting further includes the steps of:

detecting if the audio information contains a high level of energy outside the narrow frequency range of the speech signal; and

generating a second logic signal indicative of whether the signal is or is not the non-speech signal.

34. The method of claim 33 wherein:

the step of detecting whether the audio information is a speech or non-speech signal includes the step of generating said selection signal in response to a combination of the first and second logic signals; and

selecting in response to the selection signal the encoded speech or the encoded non-speech signal for transmission as a combined encoded audio signal.

35. The method of claim 31 including the steps of:

transmitting the combined encoded audio signal along with a signal identifying the digital audio information; and

receiving the combined encoded audio signal and identifying signal.

36. The method of claim 35 including the step of:

storing the combined encoded audio signal and the identifying signal for subsequent use.

37. The method of claim 36 wherein the step of receiving includes the steps of retrieving the stored signals;

separating the identifying signal from the combined encoded audio signal.

decoding the combined encoded audio signal into respective decoded speech and non-speech signals; and

selectively switching between the decoded speech and non-speech signals in response to the separated identifying signal to form a reassembled audio signal corresponding to the original digital audio information.

38. Apparatus for decoding digital audio information formed of signals such as speech signals and non-speech signals. the audio information including a signal identifying the speech and non-speech signals, comprising:

means for receiving and temporarily storing the combined speech, non-speech and identifying signals;

means retrieving the stored combined signals for separating the identifying signal from the speech and non-speech signals; and

decoder means for separately decoding the speech and non-speech signals into a re-assembled audio signal in response to the identifying signal, for audible presentation of the re-assembled audio.

39. The apparatus of claim 38 wherein the means for separating includes:

a decoder circuit for detecting the identifying signal and extracting it from the combined signals; and

soft switching means coupled to the decoder means and responsive to the identifying signal for reassembling the speech and non-speech signals for the audible presentation.

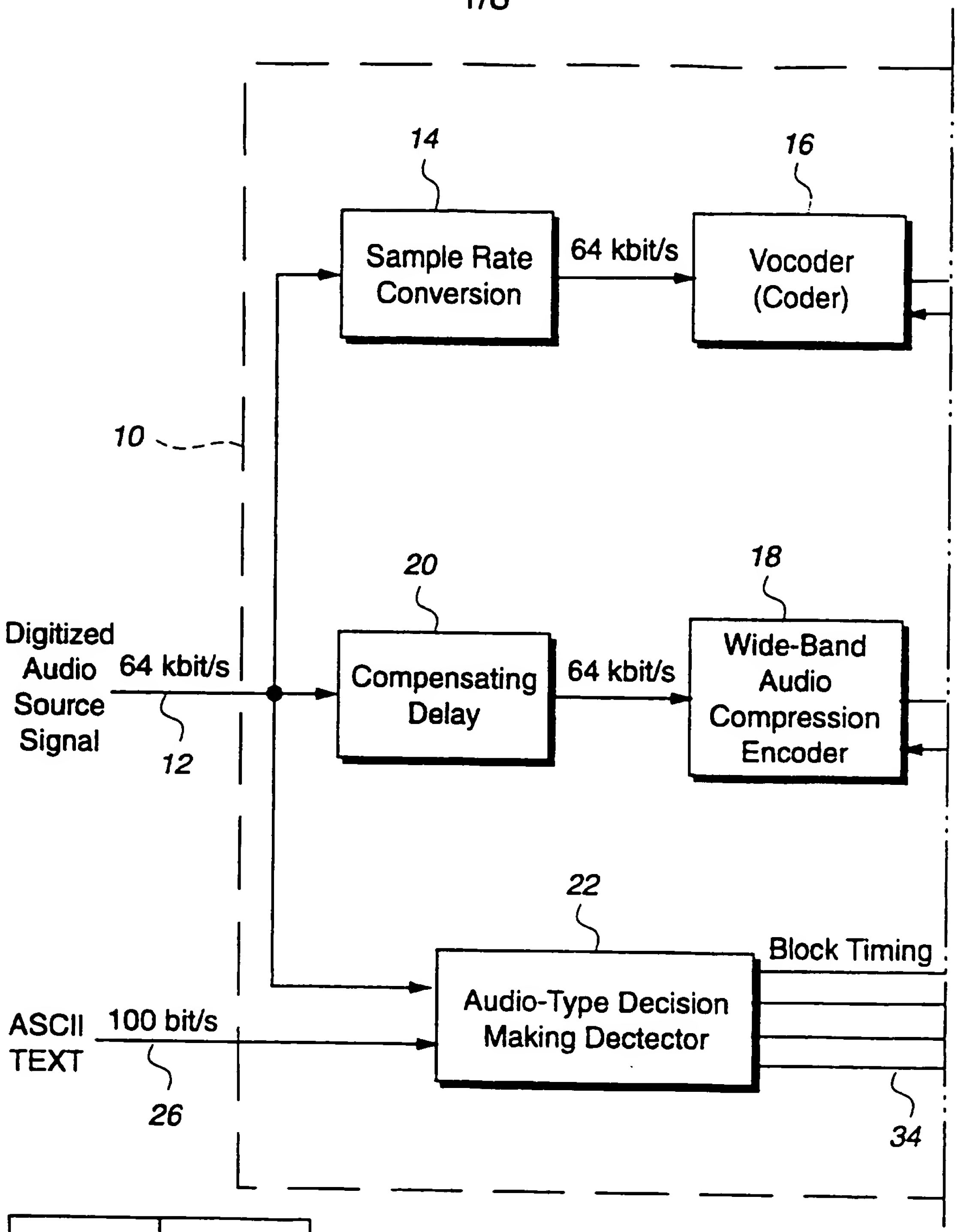

14
16
Sample Rate Conversion
64 kbit/s
Vocoder (Coder)
10
20
18
Digitized Audio Source Signal
64 kbit/s
12
Compensating Delay
64 kbit/s
Wide-Band Audio Compression Encoder
22
Block Timing
Audio-Type Decision Making Dectector
ASCII TEXT
100 bit/s
26
34

Fig. 1A

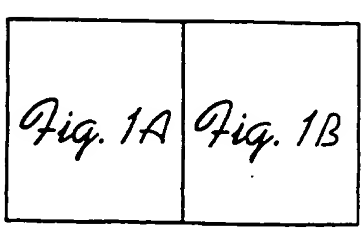


Fig. 1A
Fig. 1B

Fig. 1

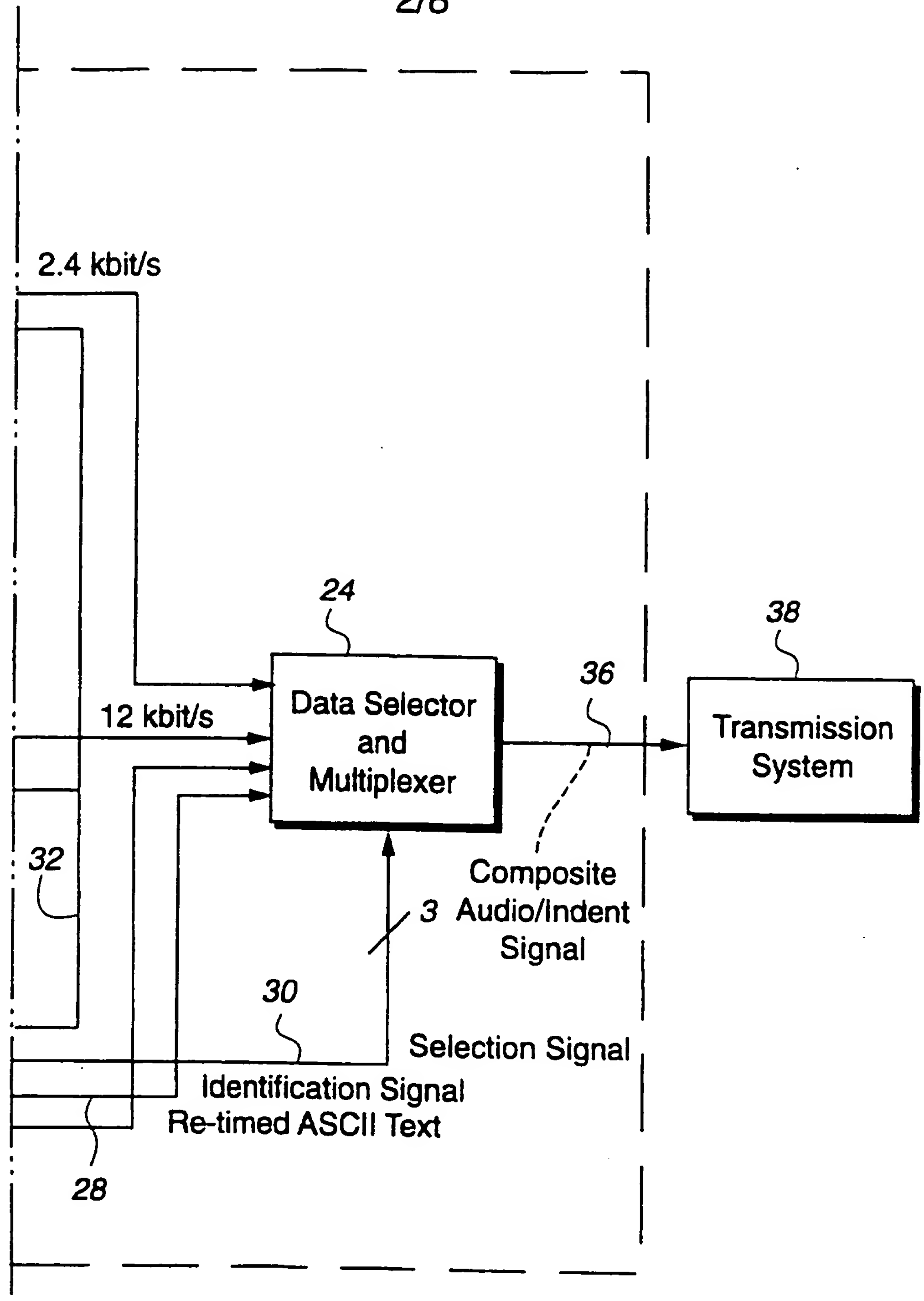

2.4 kbit/s
24
38
36
Data Selector and Multiplexer
Transmission System
12 kbit/s
32
Composite Audio/Indent Signal
3
30
Selection Signal
Identification Signal
Re-timed ASCII Text
28

Fig. 1B

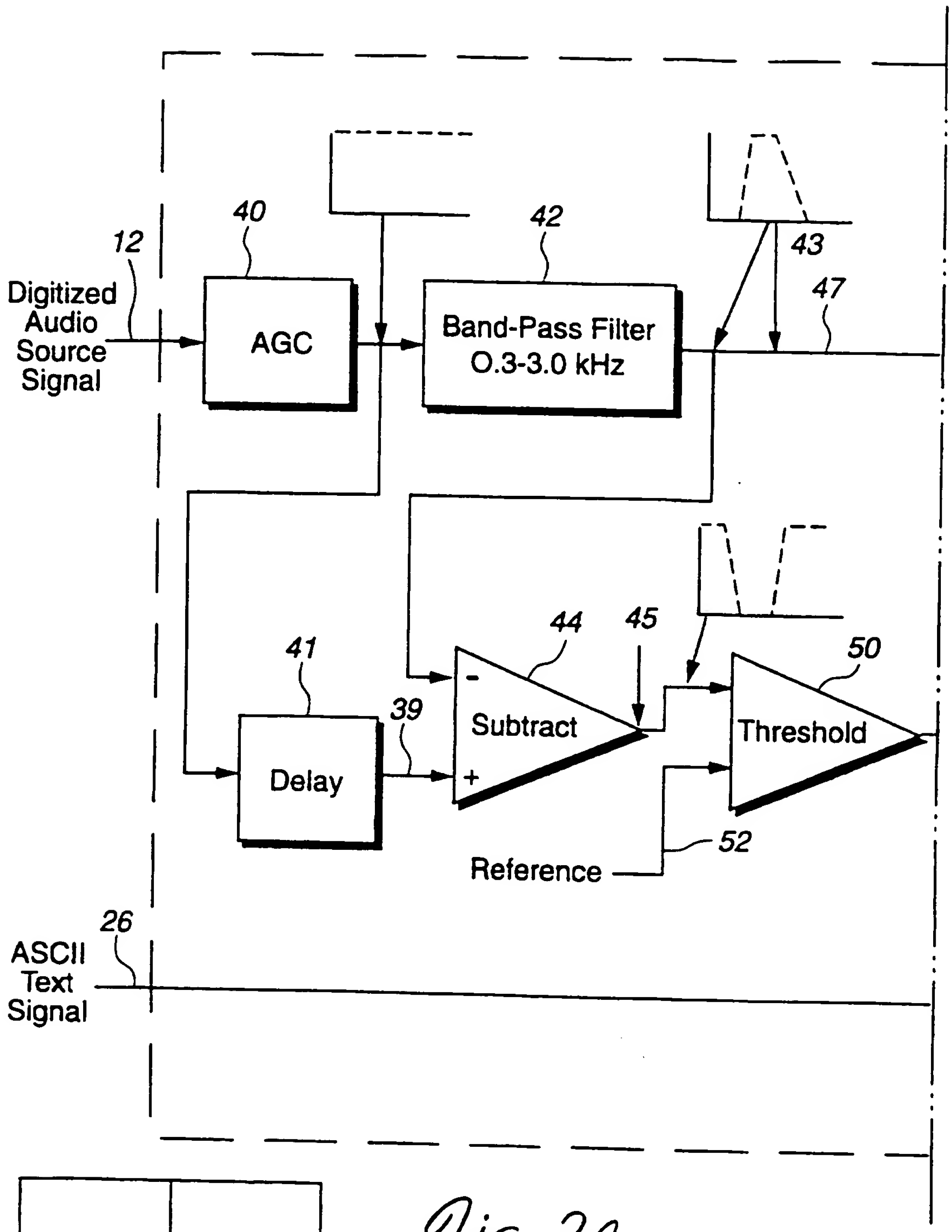

Digitized Audio Source Signal
12
40
AGC
42
Band-Pass Filter O.3-3.0 kHz
43
47
41
Delay
39
44
Subtract
45
50
Threshold
Reference
52
26
ASCII Text Signal

Fig. 2A

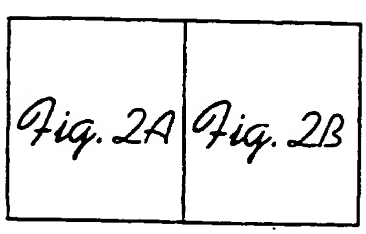


Fig. 2A
Fig. 2B

Fig. 2

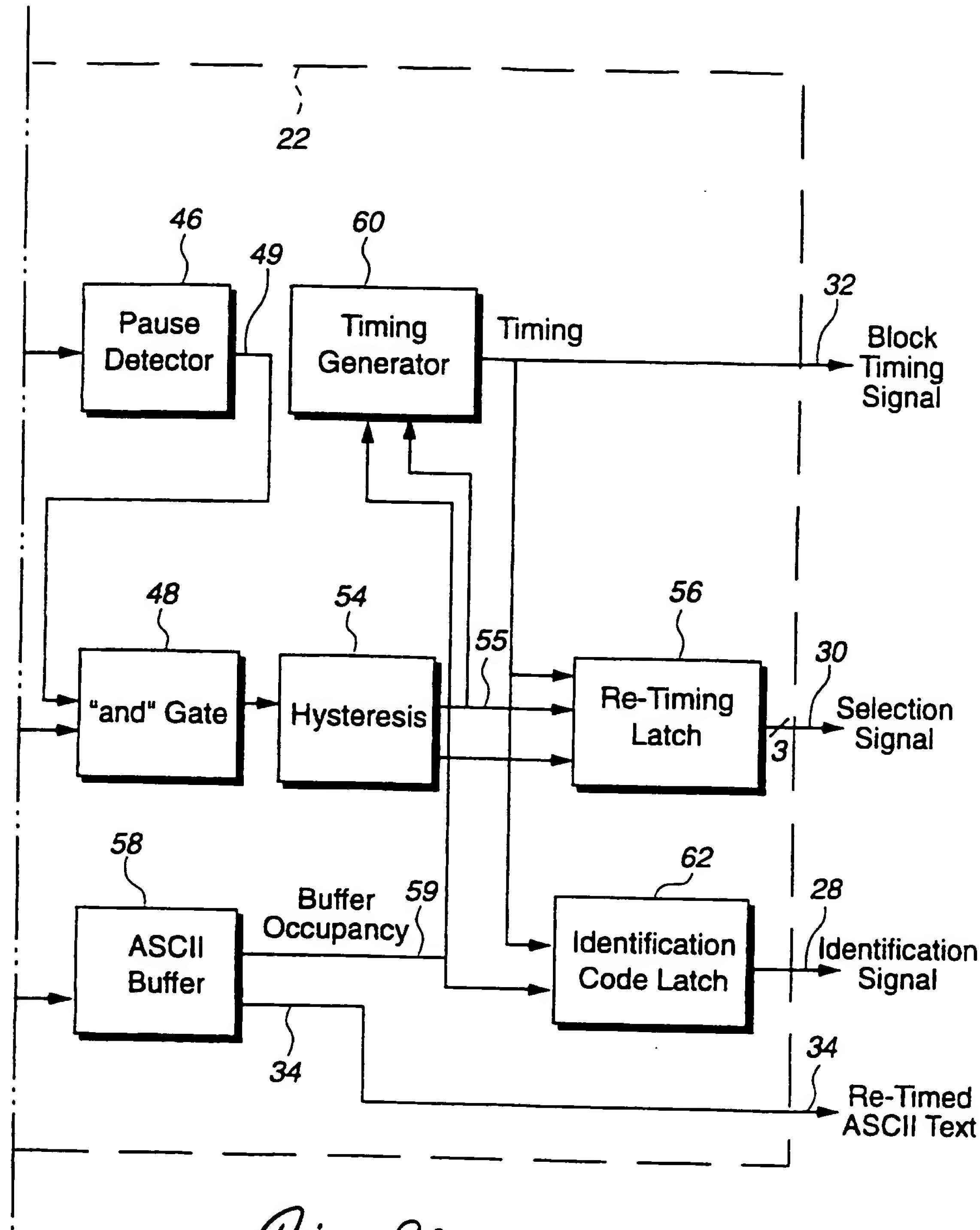

22
46
Pause Detector
49
60
Timing Generator
Timing
32
Block Timing Signal
48
"and" Gate
54
Hysteresis
55
56
Re-Timing Latch
3
30
Selection Signal
58
ASCII Buffer
Buffer Occupancy
59
62
Identification Code Latch
28
Identification Signal
34
34
Re-Timed ASCII Text

Fig. 2B

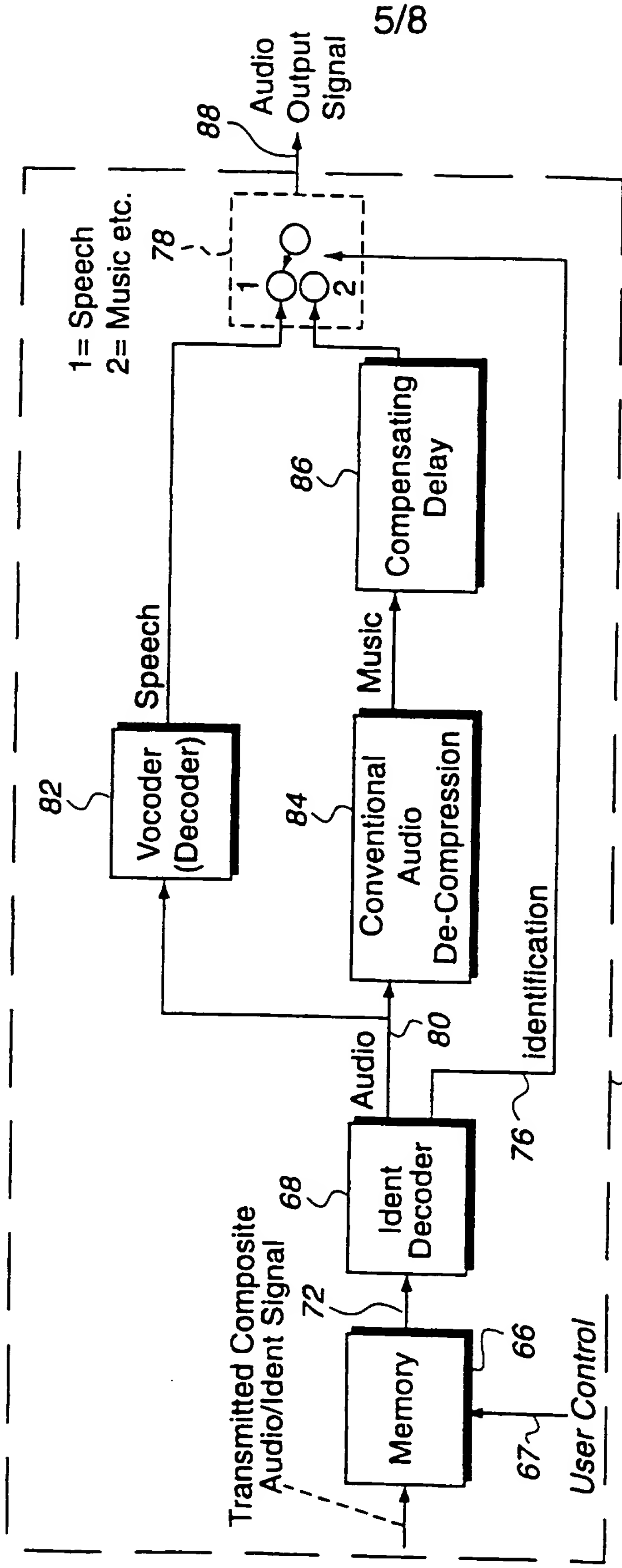

Transmitted Composite Audio/Ident Signal
Memory
66
67
User Control
72
Ident Decoder
68
Audio
80
76
identification
70
Vocoder (Decoder)
82
Speech
Conventional Audio De-Compression
84
Music
Compensating Delay
86
1= Speech
2= Music etc.
78
1
2
88
Audio Output Signal

Fig. 3

Fig. 4

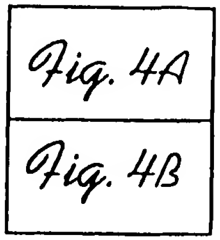
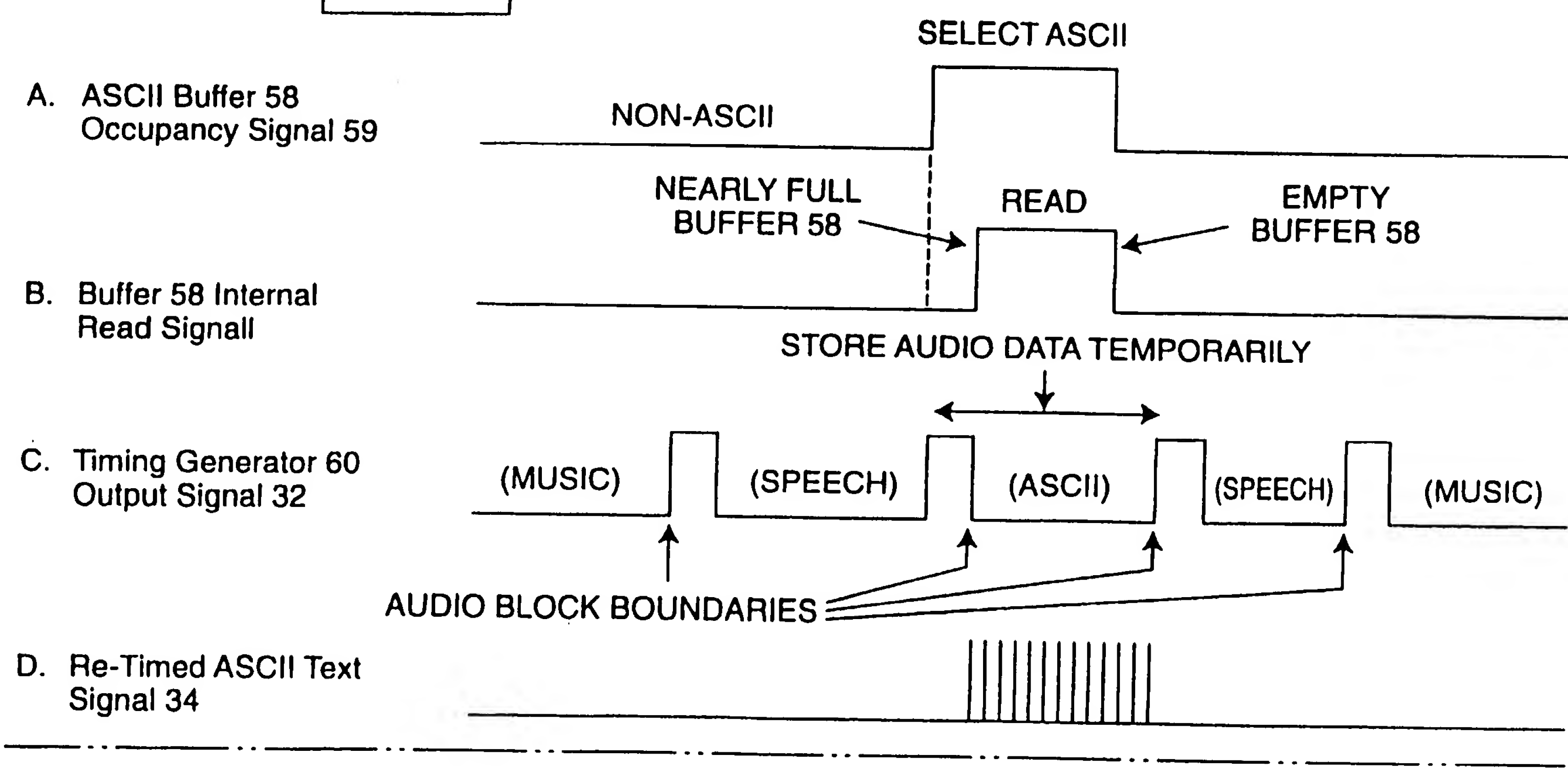


Fig. 4A
Fig. 4B

Fig. 4A


A. ASCII Buffer 58 Occupancy Signal 59
B. Buffer 58 Internal Read Signall
C. Timing Generator 60 Output Signal 32
D. Re-Timed ASCII Text Signal 34
SELECT ASCII
NON-ASCII
NEARLY FULL BUFFER 58
READ
EMPTY BUFFER 58
STORE AUDIO DATA TEMPORARILY
(MUSIC)
(SPEECH)
(ASCII)
(SPEECH)
(MUSIC)
AUDIO BLOCK BOUNDARIES

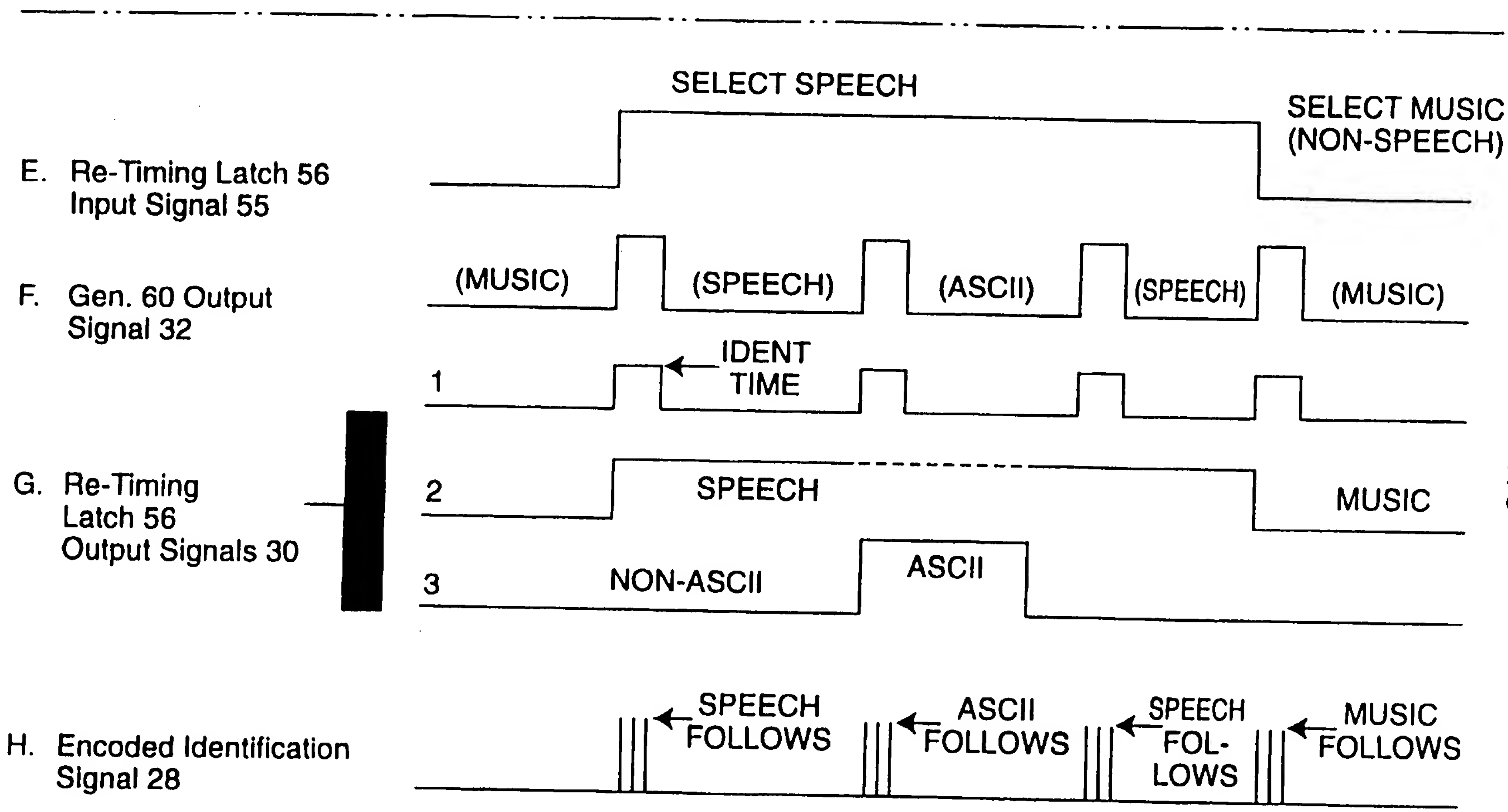

SELECT SPEECH
SELECT MUSIC (NON-SPEECH)
E. Re-Timing Latch 56 Input Signal 55
F. Gen. 60 Output Signal 32
(MUSIC)
(SPEECH)
(ASCII)
(SPEECH)
(MUSIC)
IDENT TIME
1
G. Re-Timing Latch 56 Output Signals 30
2
SPEECH
MUSIC
3
NON-ASCII
ASCII
H. Encoded Identification Signal 28
SPEECH FOLLOWS
ASCII FOLLOWS
SPEECH FOL-LOWS
MUSIC FOLLOWS

Fig. 4B

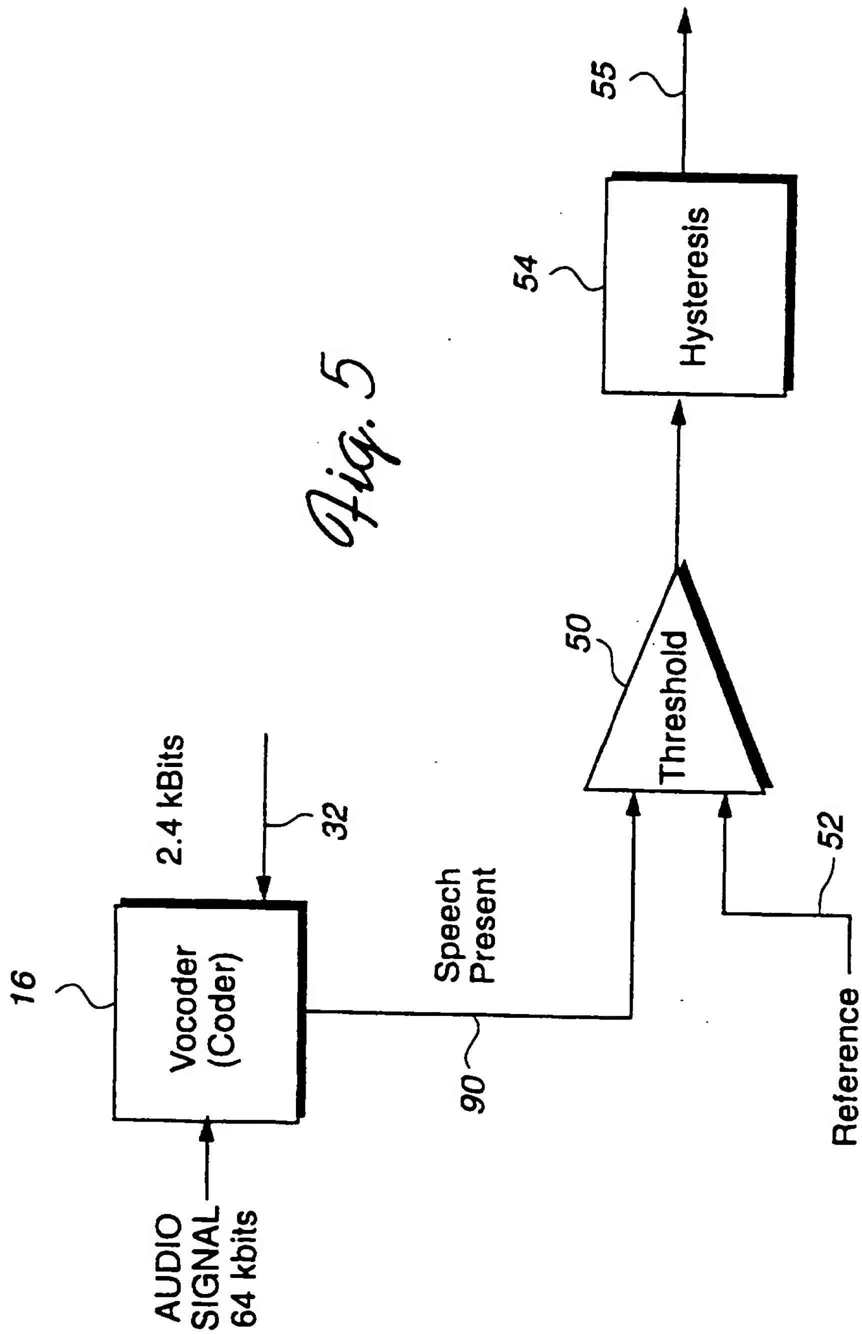

AUDIO
SIGNAL
64 kbits
16
Vocoder
(Coder)
2.4 kBits
32
Speech
Present
90
50
Threshold
52
Reference
54
Hysteresis
55

Fig. 5

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 97/05141

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 H04H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 H04H H03G H03M H04B H04J G10L H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 4 331 837 A (SOUMAGNE) 25 May 1982	1-12, 14-25, 28-32, 34-38
A	see column 1, line 5-9 see column 1, line 67 - column 3, line 44 see column 4, line 44 - column 14, line 48 ---	26,27, 33,39
Y	US 4 916 742 A (KOLESNIKOV ET AL.) 10 April 1990	1-12, 14-25, 28-32, 34-38
A	see column 1, line 9-14 see column 2, line 11 - column 3, line 9 see column 3, line 53 - column 5, line 33 see column 6, line 23 - column 11, line 2 ---	26,27, 33,39

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

1

Date of the actual completion of the international search	Date of mailing of the international search report
4 August 1997	13.08.97

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Zanti, P

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT	Internal Application No PCT/US 97/05141

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	EP 0 279 451 A (FUJITSU) 24 August 1988	1-12, 14-25, 28-32, 34-38
A	see page 3, line 8 - page 4, line 6 see page 5, line 14 - page 9, line 25	26,27, 33,39
A	US 5 467 087 A (KE-CHIANG CHU) 14 November 1995 see column 1, line 18-60 see column 4, line 24 - column 5, line 44	1,11,13, 21,28
A	US 5 444 312 A (NOBLETT ET AL.) 22 August 1995 see column 2, line 18-53 see column 4, line 42 - column 7, line 11	38,39

1

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 97/05141

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 4331837 A	25-05-82	FR 2451680 A	10-10-80
		CA 1144668 A	12-04-83
		DE 3009677 A	25-09-80
		GB 2048616 A,B	10-12-80
		JP 56060131 A	23-05-81
US 4916742 A	10-04-90	WO 8706783 A	05-11-87
		FR 2599915 A	11-12-87
		GB 2200018 A,B	20-07-88
		NL 8620347 T	01-03-88
EP 279451 A	24-08-88	JP 63204300 A	23-08-88
		CA 1301072 A	19-05-92
		DE 3883799 D	14-10-93
		DE 3883799 T	24-02-94
		US 4890325 A	26-12-89
US 5467087 A	14-11-95	NONE	
US 5444312 A	22-08-95	NONE	